



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Towards a national assessment policy in Switzerland: Areas of conflict in the use of assessment instruments

Imlig, Flavian ; Ender, Susanne

Abstract: This article reveals three emerging areas of conflict in the use of educational assessment instruments in compulsory education in Switzerland and outlines an analytical approach for detecting and analysing these areas of conflict. The approach combines a conceptual perspective, an evaluation perspective and a teaching perspective to show the different backgrounds and expectations of actors on the governmental and school levels. We apply our analysis to three assessment instruments, currently in use in Switzerland, retracing a rudimentary timeline of Swiss educational assessment, with a focus on the German-speaking region of the country. Combining the three perspectives and positioning the analytical approach within the context of political and historical developments enables us to discuss both the reasons for the conflicts and possible ways to respond to them.

DOI: <https://doi.org/10.1080/0969594x.2017.1390439>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-158612>

Journal Article

Accepted Version

Originally published at:

Imlig, Flavian; Ender, Susanne (2018). Towards a national assessment policy in Switzerland: Areas of conflict in the use of assessment instruments. *Assessment in Education: Principles, Policy and Practice*, 25(3):272-290.

DOI: <https://doi.org/10.1080/0969594x.2017.1390439>



Towards a national assessment policy in Switzerland: areas of conflict in the use of assessment instruments¹

Flavian Imlig² and Susanne Ender³

This is the post print of the article originally published in *Assessment in Education: Principles, Policy & Practice (CAIE)*, <http://dx.doi.org/10.1080/0969594X.2017.1390439>, please cite as Imlig, F. & Ender, S. (2018). *Towards a national assessment policy in Switzerland: Areas of conflict in the use of assessment instruments*. *Assessment in Education: Principles, Policy & Practice (CAIE)*.

ABSTRACT

This article reveals three emerging areas of conflict in the use of educational assessment instruments in compulsory education in Switzerland and outlines an analytical approach for detecting and analysing these areas of conflict. The approach combines a conceptual perspective, an evaluation perspective and a teaching perspective to show the different backgrounds and expectations of actors on the governmental and school levels. We apply our analysis to three assessment instruments, currently in use in Switzerland, retracing a rudimentary timeline of Swiss educational assessment, with a focus on the German-speaking region of the country. Combining the three perspectives and positioning the analytical

¹ Article history: Received 31 March 2017, Accepted 2 October 2017

² Current affiliation of corresponding author: Zurich State Department of Education, Division for Educational Planning, Zurich, Switzerland. flavian.imlig@bi.zh.ch, <http://orcid.org/0000-0002-6305-8088>

³ Institute for Educational Evaluation, Associated Institute of the University of Zurich, Zurich, Switzerland. susanne.ender@ibe.uzh.ch, <http://orcid.org/0000-0002-7841-9114>



approach within the context of political and historical developments enables us to discuss both the reasons for the conflicts and possible ways to respond to them.

Keywords: Assessment instruments; education policy; analytical approach; educational assessment; Switzerland

Introduction

The educational assessment of students' performance has become a focal point of school governance and development and a possible way to support teaching and learning. As education policies have again become a 'hot topic on the political agenda', comparative assessment results are being intensely scrutinised in both the public and political spheres (Jakobi, Martens, & Wolf, 2010, p. 1). In its basic form, we understand educational assessment as any form of graded or evaluated work in schools, such as written or oral exams, essays and projects. Such assessment instruments are integral to the grading and evaluation processes that allow judgements on students' learning and achievement (Sun & Cheng, 2014). Several disciplines have explored the topic of educational assessment: Test developers are working to measure competences in more valid, reliable and objective ways (Lane, Raymond, & Haladyna, 2016); policy analysts are focusing on how international assessment studies shape governance mechanisms and power relations in the educational field (Martens, 2010); and educational scientists are analysing how instruments are or should be used in the teaching and learning process (Gordon & Rajagopalan, 2015).

In Switzerland, the assessment of students' performance has long been treated as a matter of instruction and teaching. Especially in compulsory education, assessment has been seen as a necessary tool for schools to fulfil their purpose of allocating and selecting students into different education paths. This purpose is, as Fend (2008) points out, one of



the socially encoded functions of schools in society. In addition, assessment has been seen as a general core feature of schooling with high universal validity and stability over time (Tyack & Tobin, 1994). Beginning in the 1980s, Swiss teachers and politicians began to shift their judgement practices towards more scientific and standardised approaches, establishing a new culture of assessment within which assessment was discussed as a feature of educational and instructional quality (Vögeli-Mantovani, 1999). As a result, a wide range of development projects focusing on innovation and quality in compulsory schooling emerged all over Switzerland (e.g. Ambühl et al., 1986). In terms of assessment, relevant projects and tools were often developed locally and focused on actual educational practice in schools.

Since 2000, international discourse and, in particular, international large-scale assessments have again changed the perceptions and usage of information on students' performance. Throughout Switzerland, the use of such information in education-related policy-making has become the focus of intense discussion (Criblez, 2008a; Herzog, 2008). As a result, educational assessment has increasingly become seen as important beyond the levels of actual practice or single teachers or classes. Over time, new assessment instruments have become more and more connected to specific modes of monitoring, controlling and reporting. On a national level, Switzerland has introduced an educational governance system using standardised monitoring and reporting tools (Wolter, 2008). In 2016, the implementation of a nationwide performance assessment called 'Evaluation of basic competencies' (ÜGK) complemented this national assessment policy (Weber, 2016). As a whole, though educational assessment development in Switzerland started with locally implemented instruments for teaching and instruction, since 1990, assessment instruments have evolved towards systematic monitoring and increasing



diffusion, culminating in a national performance assessment policy. However, more recent instruments did not replace existing ones. Teachers, schools and even cantons continued to develop and use various sorts of local assessment instruments.

In this article, we analyse the development of Switzerland's assessment policy from the bottom up. We begin by exploring the political and historical contexts of educational assessment in Switzerland. This first section presents a short overview of the principles of educational governance in Switzerland, a historical outline of the national reception of international discourses about indicators and large-scale assessments and a description of the subsequent shift towards evidence-based policy in Switzerland. In the second section, we present our analytical approach, which combines a conceptual perspective, an evaluation perspective and a teaching perspective. We then present in the third section three instruments with different aims and scopes, two of which are currently used in German-speaking Switzerland (Orientation test, Stellwerk test) and the 'Evaluation of Basic Competencies' test (ÜGK) which is now under implementation in the entire country (German-, French- and Italian-speaking regions). By applying our approach in the fourth section, we discuss three areas of conflict in the use of the three assessment instruments. Different opinions about the purposes of the instruments, the connection of aggregation levels to the sovereignty over test results, and the potential influence of assessment instruments on instruction bear conflicts that are revealed in our discussion. The conclusion explores the implications of the identified conflicts and summarises what needs to be taken into account in the processes of conceptualising, developing and implementing educational assessment instruments.



Political and historical contexts

The assessment of students' performance in Switzerland is situated in a threefold context both historically and politically. First, assessment policies are embedded in a particular landscape of education polity and governance within a federal, multilevel system of diverse actors and stakeholders. Moreover, assessment is thematically intertwined with Swiss policy traditions and international trends concerning both educational indicators and evidence-based policy-making. All three of these contexts shape the concepts, development processes and implementation of actual assessment instruments.

Educational governance in Switzerland

As a small state with a decentralised educational tradition and highly consensus-oriented policy institutions at all levels, federalist Switzerland is a special case when it comes to education policy-making. Over the last decade, the education systems of the federated states (cantons) have undergone a wide range of transformation processes, both structurally and in terms of governance (Bieber, 2010). Changes in school principles, systems of quality assurance and control, education monitoring and reporting and education standards can be observed at both the national and the canton level (Maag Merki & Büeler, 2002).

Like in other federal countries, compulsory education in Switzerland has been a primary policy field through which cantons have cultivated their autonomy throughout the twentieth century. At the same time, policies across the 26 Swiss cantons have moved towards a consensus, and cooperation has been maintained. The cantonal education departments coordinate at the intercanton level via the Swiss Conference of Cantonal Ministers of Education (EDK). This harmonisation can be seen as a reaction to growing



demands for education and mobility. In addition, cantonal education departments strive to coordinate their education policies in order to preserve their decentralised structure and authority in the education sector (Criblez, 2008c; Hega, 2000).

Despite this decentralised governance configuration, beginning in the 1980s, educational governance became a widely discussed national topic. Responding to international developments, such as the shift towards decentralised decision-making in the European Union (Green, 2002), most cantons began to move towards school autonomy. This development strengthened individual schools, fostered educational innovation from the bottom up and introduced new mechanisms of coordination and accountability between educational practice and educational policy (Maag Merki & Büeler, 2002; Nussbaum, Fischer, & Hildbrand, 2007). Concerning structural reforms on a national scale, the federal structure and high number of possible veto players created a backlog of unresolved policy issues that lasted well into the 1990s (Bieber, 2010).

From indicators to competence-based large-scale assessments

On an international level, educational assessment is closely intertwined with popular largescale assessments like PISA (Programme for International Student Assessment), TIMMS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study), ALL (Adult Literacy and Life Skills Survey) and IALS (International Adult Literacy Survey). In Switzerland and many other countries, these large-scale assessments frame a majority of the policies surrounding assessment (Tillmann, Dederling, Kneuper, Kuhlmann, & Nessel, 2008; Windzio, Knodel, & Martens, 2014).

Since the first attempts of the IEA (International Association for the Evaluation of Educational Achievement) to survey a so-called ‘attained curriculum’ in the 1980s



(Pelgrum, 1986, p. 6), the comparative assessment of students' achievement has been connected to the internationally shared aim of an indicator-based, comparative view of national education systems. At the same time, the OECD (Organisation for Economic Co-operation and Development) relaunched its programme of education indicators (Papadopoulos, 1994/1996; Tröhler, 2013). In the late 1980s, the OECD adopted the concept of monitoring output using students' performance and began the development of the later PISA. These PISA assessment data completed the OECD's neoliberally inspired concept of holistic and internationally comparable descriptions of education systems (Davies, Nutley, & Smith, 2012; Martens & Wolf, 2006; Sjøberg, 2007; Uljens, 2007).

The international large-scale assessments also (re)introduced paradigms of competence- or skill-based teaching and learning in compulsory schooling (Tyo, 2010). The concept of competence was originally introduced as an alternative to intelligence and was meant to describe a holistic capacity for reasonable and responsible action (McClelland, 1973). In international large-scale assessments, the notion shifted towards the measurability, assessment and evaluation of educational performance, often in relation to given standards (Oelkers & Reusser, 2008).

The Swiss Federal Statistical Office participated in the development of education indicators and published its first report according to OECD definitions in 1992 (BFS, 1992). In 2002, the first Swiss PISA results fell 'on fertile ground' and triggered fundamental discussions on the capacity and efficiency of schooling (Bieber, 2014, p. 186). The newness of this type of performance information and the backlog of structural reforms in federal Switzerland were two reasons for this fundamental impact. The PISA results were followed by a wide range of reforms, including, most recently, steps towards harmonising the cantons' education systems (Bieber, 2010; Criblez, 2008b; EDK, 2011).



These harmonisation processes have also involved the introduction of national monitoring and reporting (Wolter, 2008).

The concept of competence has also played an important role in the implementation of educational standards on both national and regional levels (Criblez et al., 2009). It has been a critical concept in educational practice due to its influence on textbook development and teacher training. Competence-based assessment tools, as well as instruments for competence-based teaching and instruction, have been introduced. (Bölsterli Bardy, 2015; Larcher & Smit, 2011).

Towards evidence-based policy

Ideas on evidence-based policy and practice have grown in popularity across all policy fields in many countries (Biesta, 2010, p. 492). This shift has been accompanied by changes to the guiding principles concerning polity, state organisation and public action (Jann & Wegrich, 2010), including the shift in regulatory responsibilities to international organisations and the shared provision of public goods (Hurrelmann, Leibfried, Martens, & Mayer, 2007). In education, the idea of evidence-based policy-making has led to widespread accountability reforms (Cibulka, 1990). A key component of these governance regimes, often referred to as ‘new public management’ (NPM), is the use of educational assessment (Green, 2011; Mitchell, Shipps, & Crowson, 2011).

In Switzerland, NPM concepts gained increasing popularity throughout the 1990s (Rieder, 2005). Most NPM-inspired reforms focused on management, supervision and accountability, leading to the introduction of head teachers and quality management procedures in most cantons (Hangartner & Svaton, 2013). The use of comparative data on educational outcomes has been intensely discussed in Switzerland. Unlike the UK or the Netherlands, Switzerland’s political and scientific landscape has been characterised



by critical positions on high-stakes testing, performance ranking and quasi-market models of education (Criblez, 2008a; Green, 2002; Herzog, 2008).

By positioning educational assessment within its complex and multifaceted historical and political contexts, we seek to set the stage for our analysis, which approaches actual assessment instruments via the three distinct analytical perspectives.

Analytical perspectives

Assessment in education is the subject of extensive research. Berry and Adamson's (2011) inventory of assessment reforms follows a political approach, which can also be found in the research on international assessments and how they shape national education policy (Davies et al., 2013; Martens, Knodel, & Windzio, 2014; Peyrera et al., 2011). In critical policy analysis, traditional concepts of functionalism and rationalism have been rearranged to reveal policy contexts, traditions and overlooked actors and to introduce theorising methodologies and qualitative approaches to the field (Young & Diem, 2017). From a more technical perspective, there is a broad discourse on assessment quality, measurement accuracy and evaluation implementation (Goldstein, 2015; Lane et al., 2016). Purposes and quality criteria are key research subjects in the growing modern international assessment landscape (Broadfoot & Black, 2010). A relevant portion of this research also deals with the implications of assessment for educational practice (Gordon & Rajagopalan, 2015).

Our analytical approach seeks to integrate different aspects of this theoretical and empirical research on educational assessment. We seek to examine how educational assessment is embedded in education policy and education. To gain a coherent and plausible analysis of assessment instruments, we integrate three perspectives: a



conceptual, an evaluation and a teaching perspective. This approach is not meant to fully cover all aspects of assessment tests, since there are, of course, topics that demand an extension of the analytical approach. The three perspectives allow us to discuss the broad questions surrounding the conceptual premises of assessment instruments, their evaluation policies and their practical impact on teaching, taking into account the complex interactions among educational governance, international influences and the adoption of evidence-based policy in Switzerland.

Conceptual perspective

From a conceptual perspective, student performance assessments seek to fulfil either formative or summative purposes. Formative assessments, sometimes referred to as assessments for learning, emphasise the connections between assessment and learning (Broadfoot & Black, 2010). From a formative perspective, assessment is an integral part of the learning process. It is a tool for communicating learning among actors at the level of educational practice (Ambühl et al., 1986). By contrast, summative assessments are used in communications related to classes, schools and education systems (Vögeli-Mantovani, 1999). They are meant to be public, and they relate to societal functions of allocation and selection and serve purposes of certification and accountability. Formative and summative assessments are just two of a bigger set of purposes for assessment instruments. Though they share common characteristics, the underlying assumptions and methods used by test developers differ (Harlen, 2012; Yates & Johnston, 2017). It is often argued that there are no good combinations of the two purposes and that summative assessments undermine the efficacy of formative assessments (e.g. Harlen & James, 1997). However, some evidence shows that multiple perspectives on purposes should be considered simultaneously and consciously during the conception of an assessment



instrument (Newton, 2017). If summative assessments are combined with high stakes, then authorities will put stronger pressure on both teachers and students, since the results will be crucial for the course of education or the teacher's career. In using such tests, teachers face a conflict between accountability and their responsibility to the learning process. Therefore, the opposition between summative and formative assessment must be overcome, and teachers must be supported in adopting assessment for learning practice instead (Black, 2015).

From a conceptual perspective, we investigate the tensions between 'purpose purism' and 'purpose pluralism' (Newton, 2017) within the concepts and guidelines of assessment instruments in Switzerland. How are assessment instruments positioned in the interplay of different purposes and their respective assumptions? Which demarcations and boundaries in terms of functions or purposes can be drawn?

Evaluation perspective

Students' performance assessments are always directly connected to the level of educational practice. The results of single students are evaluated in either a prognostic or recapitulatory manner in order to look forward or backward in time. The evaluation perspective is strongly connected not only to educational practice, but also the political foundations of the school system. Educational governance points to a wide range of actors and their reciprocal influences of a multilevel system (Böttcher, 2007). The level of educational practice involves schools, teachers, students and parents. Above this level are multiple policy levels, ranging from school administrators to administrators of cantons, groups of cantons and the Swiss federation. The descriptions of educational governance in a multilevel system are based on sociological theories, such as neo-institutionalism and system and organisation theory (Berkemeyer, 2010; Koch & Schemmann, 2009). The



trend towards evidence-based policy presented above establishes connections among actors on different levels. The evaluation of the results of educational assessment links the different levels by extending the results to levels above single students. Information can also be aggregated beyond the level of single students' performance, extending to the levels of classes, schools, regions, nations and other entities.

Evaluation theory differentiates among the various evaluation levels of educational assessment (Rhyn, 2009). For example, a teacher assesses the performance of his or her students, possibly compared to a broader standard. He or she then uses the resulting information to judge the students' performance and to reflect on his or her practice. At the next level, an educational organisation, typically a school, assesses the students' performance in order to gather information on quality and fulfil responsibilities for reporting and accountability. Finally, an education system (in Switzerland, typically a canton) assesses the students' performance to legitimate itself and inform processes of educational governance. When the results of educational assessment are projected to the system level on a regular basis, this is called education monitoring (Hovenga & Bos, 2009). At the level of the educational system, information on students' performance is often combined with other data indicators to support political conclusions (Wolter, 2008).

From an evaluation perspective, we investigate the levels of aggregation of performance data in relation to the political purposes of assessment instruments in Switzerland. Which actors use the assessment results? Which projections of students' performance can be observed?

Teaching perspective

The teaching perspective focuses on the instructional relevance of educational



assessments, since forms and procedures of assessments express general understandings of teaching and learning (Vögeli-Mantovani, 1999). As presented above, educational assessments, especially popular large-scale assessments, imply competence-based teaching and learning. This concept follows a constructivist understanding that organises learning and teaching arrangements around individual students' learning processes. Therefore, teaching consists of different actions designed to initiate, support, coach, scaffold, review and consult students' learning. These actions are combined with an analytical, research-based teaching methodology designed to support the usability and applicability of learning outcomes (Wiater, 2013). A specific understanding of teaching and learning is also supported by the content of assessment instruments. The actual tasks represent certain underlying concepts of learning and play an important role in instructional processes (Drüke-Noe, 2014), since they structure teaching, learning and results (Knudson, 1993). Therefore, in recent educational assessments, task types have been important at the level of education practice. Especially in Switzerland, actors on the level of educational practice often raise concerns regarding 'teaching to the test'. It is assumed that teachers intuitively align their teaching towards various assessments, especially if the stakes are high (Yates & Johnston, 2017).

From a teaching perspective, we investigate the potential influences of assessment instruments and their included tasks on instruction. Which elements are explicitly and implicitly transported into the educational practice?

Instruments under scrutiny

To apply our analytical approach, we chose three assessment instruments that represent a rudimentary timeline of educational assessment in Switzerland, with a focus on the



experience of the German-speaking cantons. Although teachers use many ‘handmade’ tests to evaluate the performance of their classes, only a few instruments exist that claim to either completely or partially fulfil the psychometrical criteria of valid, reliable and objective assessment.

One of the first such instruments, which is still used in the German-speaking part of Switzerland, is the ‘Orientation test’ (*Orientierungsarbeit*). The first iteration of this test was developed and published in 1994 by the canton of Lucerne and comprised a set of standardised mathematics questions for sixth graders. Since then, five other cantons have joined the project, and the tests now cover a wide range of subjects for grades 2–9 (BKZ, 2013). The teachers’ use of the Orientation tests is regulated by cantonal guidelines (e.g. Frey, 2010). Orientation tests are paper-and-pencil tests that are published as brochures and resemble the types of tests teachers use in their classes. They are tailored to the current curriculum (Sutermeister-Christen et al., 2007). Though the evaluation of students’ answers is neither standardised nor centralised, the Orientation tests contribute to a more objective grading practice and refer to external criteria (rather than individual teachers’ instruction). Thus, they tackle the problem of teacher bias in class-oriented grading (Vögeli-Mantovani, 1999).

In 2006, a decade after the first release of the Orientation test and influenced by psychometrically shaped assessments like PISA, the ‘Stellwerk test’ was developed and implemented by the canton of St. Gallen. Since the Stellwerk test has been implemented by nearly all German-speaking cantons, its items are based on common educational goals of the cantons’ different curricula (Wolter & Hof, 2014). The Stellwerk test is a computer-based test with items referring to Question and Test Interoperability (QTI) specifications, such as multiple-choice, short answer and drag-and-drop. Both the test development and



the evaluation of the students' results are centralised and conducted by a professional organisation. The test seeks to build an individual profile of competencies of students in the eighth and ninth grades in order to prepare them for their transition from compulsory school to upper secondary school and/or vocational education and training (Staatskanzlei SG, 2006).

Most recently, in 2016, a decade after the first release of the Stellwerk test, the 'Evaluation of Basic Competencies' (ÜGK) assessment of the Swiss educational system was implemented. This assessment is rooted in the responsibilities shared by the federation and the cantons. It was developed by the EDK and is part of the national education monitoring strategy. It comprises a sample-based assessment of the competencies of second, sixth and ninth graders throughout Switzerland. In its final implementation, it will cover mathematics, science and both first and foreign languages. The ÜGK is a computer-based assessment that uses QTI item formats. It is designed to measure whether students have attained national education standards on both the national and cantonal levels (EDK, 2013).

Discussion: areas of conflict

The different aims and scopes of the described assessment instruments raise questions concerning their potential for governance, the availability of their data and the use of their results. Applying our analytical approach to the three assessment instruments described above reveals three areas of conflict in regard to these questions: from a conceptual perspective, there is a confusion of purposes; from an evaluation perspective, the different aggregation levels give rise to questions of sovereignty; and from a teaching perspective, the influence of assessment instruments on instruction is unclear. We seek to show how



the specific political and historical contexts in Switzerland, as presented in the first section, produce conflicting ideas of educational assessment.

Confusion of purposes

We discuss the purposes of the three assessment instruments from a conceptual perspective. The instruments deal in a specific way with the distinction between formative and summative purposes and processes (Harlen, 2005). The ways in which these instruments are positioned between these purposes reveal conflicts.

The original aim of the Orientation tests was clearly and explicitly defined as summarising students' knowledge and skills in relation to the goals of the curriculum. In this respect, the tests were designed to make teachers' judgments more objective (Vögeli-Mantovani, 1999). Nevertheless, the first Orientation test in 1994 explicitly sought to formatively support teachers in planning individual support and teaching (Jost, 1994). This mix of functions was seen as a problem by governmental actors, especially because the assessment was positioned at the end of primary school, when students were assigned to a certain track of secondary school. On one hand, the 'results of the 'Orientation tests' should be used neither to give marks nor to justify assignment decisions', while, on the other hand, 'the results provide criteria for assignments to continuing schools' (Jost, 1994, p. 1). From the beginning, the guidelines and discussions surrounding the tests indicate a confusion of purposes. When the Orientation tests were expanded, the originally unintentional mix of formative and summative purposes persisted. Today, both the planning function and the performance function are highlighted in parallel (Frey, 2010). The educational administration is not clear in its communication of the purpose. In the canton of Lucerne, for example, the formative function of evaluating individual performance is clearly emphasised and set in contrast to the functions of other



standardised and summative assessment tests (BKD LU, 2013). On the other hand, the instrument is also used during the process of assigning students to lower secondary school (Roos, Wandeler, & Mosimann, 2013).

Since the beginning of the Stellwerk concept, there has been an attempt to avoid a confusion of purposes by clearly distinguishing summative and formative purposes. The first assessment, which is given to eighth graders, is meant to be formative, while the second one, which is given to 9th graders, is designed to be summative. The latter is meant to measure students' skills at the end of compulsory school, while the former is seen as a planning tool for the last year of compulsory school (Staatskanzlei SG, 2006). In actual practice, this two-step concept has seen little realisation. The eighth-grade test is used mainly as a certificate to apply for vocational education, and the formative function takes secondary importance (Goetze, Denzler, & Wissler, 2009). This change in the purpose of the assessment test has had a backwash effect, such that official guidelines issued by educational administrations that originally argued for a distinction of purposes now recommend that the eighth-grade test be used in a summative manner (e.g. BD SZ, 2015).

The ÜGK is based on a different concept than the Orientation test and the Stellwerk test. It has a clear summative function: It is meant to give information on the national and cantonal levels regarding whether students have achieved national education standards. Its results inform educational policy on the performance of the educational system with respect to education standards. The ÜGK does not aim to evaluate single students, schools or teachers (EDK, 2013). In fact, due to the assessment's sample-based approach and references to national education standards, a formative purpose is virtually impossible (Klausing & Husfeldt, 2015). However, though the concept of the evaluation seems to be distinctively summative, its connection to national education standards also



suggests a formative aspect. The evaluation of the achievement of educational goals implies conclusions designed to support the development of educational system quality (EDK, 2015). Though the responsibility and processes for achieving such conclusions have not yet been defined, drawing development goals from performance data means going beyond using the evaluation for exclusively summative interpretations. In sum, the ÜGK reinterprets the traditional formative purpose of assessment instruments by relocating the responsibility of ‘looking forward’ to improve students’ performance to policy-makers.

Aggregation levels and sovereignty

The needs of different actors (e.g. teachers, schools, cantons) and their expectations concerning the effects of various instruments are reflected by the instruments’ levels of evaluation. As Goldstein (2004) pointed out with respect to PISA, there is a mismatch between the conceptual restrictions and the wide political use of the instrument. The ways in which the results of the three investigated assessment tests are evaluated reveal political conflicts.

The Orientation tests were developed from the bottom up in the context of instructional quality development. They were meant to support teachers’ assessments of students’ performance; therefore, they were orientated towards the processes of teaching and learning (Vögeli-Mantovani, 1996). The Orientation tests have also faced claims concerning aspects of their role as standardised assessments, such as the valid operationalisation of performance and the collection of context factors of instruction and learning techniques. Nevertheless, the tests remained bound to instruction. The tests’ decentralised ways of evaluating students’ performance also prevent the aggregation of results (BKD LU, 2013). Some cantons use a monitoring mechanism to supervise the



Orientation tests, but not to aggregate their results (e.g. BD NW, 2015). The Orientation tests refer to the instructional level of education and are strongly connected to both individual teachers and their classes. They are not standardised in a way that allows comparisons across all kinds of classes, schools or even cantons. Misunderstandings of the possibilities of data aggregations can create political conflicts. Specifically, when Orientation tests tend to be used by communal or cantonal policymakers in a comparative way, teachers become pressured by the accountability assigned to this originally instructional instrument. There is a conflict between educational practitioners and superordinate governing levels concerning sovereignty over the test and the right to use the produced information for their own purposes.

The Stellwerk test supports the aggregation of performance results across different levels, beginning with the individual student and going up to the level of the canton. In several Swiss cantons, the educational administration both prescribes and funds the test. There, student performance data are aggregated on four levels. Students receive profiles of their individual performance, teachers receive profiles of their classes in comparison to the cantonal standard, school boards receive profiles of their schools in comparison to the cantonal standard and cantons receive detailed reports showing anonymised differences among classes and schools (e.g. BD SZ, 2015). These evaluation practices illustrate the diminishing importance of the formative purpose of the eighth graders' test and imply accountability mechanisms that go beyond mere classroom instruction. The process of projecting assessment information to not only the levels of learning and instruction, but also the levels of schools and the cantonal education policy field opens up an area of political conflict regarding the sovereignty of the test and the use of its results. The results of the eighth graders' test are presented in a way that fosters



their use at levels other than teaching and learning. Students use their test results for applications (a usage recommended by officials), and administrators are informed about the results of classes, schools and cantons.

The ÜGK explicitly excludes all evaluation levels below the cantons. No reporting will be made at lower levels, such as schools, classes, teachers or individual students (EDK, 2013). The results target the education system level and are made available to relevant research. The main evaluation level is that of the cantons. At the moment, there is no clear information concerning which data are provided and how they can be used. The cantons receive evaluation results concerning their own cantonal performance in relation to basic national competencies (Klausing & Husfeldt, 2015). Though all evaluation levels below the cantons' education systems are excluded, political conflicts arise from questions of responsibility and aggregation among cantons at the federal level. This conflict area specifically involves the Swiss governmental system. Though the cantons are in charge of their own education systems, they are requested by federal law to harmonise them. The EDK is a key actor concerning harmonisation at the intercantonal political level. Since the results of the ÜGK serve as an indicator of harmonisation, it is unclear which political level (i.e. the cantons, the EDK or the federal state) is responsible in the event that results do not match expectations. Finally, the inclusion of independent research in the evaluation process raises questions of autonomy and responsibility. The researchers engaging in evaluations of the ÜGK tend to come from Swiss teacher training colleges and universities. On one hand, they are committed to independent research; however, on the other, they are part of the educational system, not least because they are responsible for teacher training. Thus, they may contribute simultaneously to both problems and solutions, creating a challenging starting position



for research. The ÜGK is a politically governed educational assessment that is meant to support evidence-based policy. Therefore, the extent to which policymakers do or might hand over responsibility to supposedly independent researchers is currently unclear.

Influence on instruction

Current assessment instruments seek to inspire not only teachers' evaluations of students' performance, but also new methods of instruction. Through the nature of their tasks, assessment instruments transfer instructional elements into classes and interact with other teaching elements. The ways in which instruments are used in educational practice reveal instructional conflicts shaped by traditions and trends in teaching and learning. From a teaching perspective, it is unclear whether teaching and learning can contribute to educational steering policy or should be left unaffected by superordinate political aims (Maier, 2015). The discussions around teaching and testing during the implementation of the Orientation tests and the Stellwerk test in Switzerland represent areas of conflict that can be applied to other assessment instruments as well. Since these two were the first widely used instruments in German-speaking Switzerland, the conflicts can be seen as prototypical harbingers of the broad discussion today.

The Orientation tests are bound to their formative function and their close connection to teaching and instruction. Among other measures, these tests contributed to the 1990s reforms in teaching and instruction by introducing multidimensional, complex and challenging tasks (Vögeli-Mantovani, 1996). Several cantons' guidelines stress the model character of the tasks included in the Orientation tests (e.g. BKD LU, 2013), which are designed to be used as samples for teachers to convert and modulate (e.g. BKD OW, 2013). Educational conflicts emerge from the blurring of the frontiers between the tests and the instructional material. With the Orientation tests, the risk is that the tests could



become the teaching, which would ultimately make the tests needless as assessment instruments. If instruction is too closely connected to the Orientation tests, then the tests will no longer be able to reflect students' performance from an external perspective. The strength of the Orientation test is that it challenges students to show their skills in new tasks. When incorporated into instruction, the Orientation tests can no longer maintain their objective position.

When the Stellwerk test was first introduced into schools, the main challenge was to acquaint teachers and school boards with the concept of a computerised assessment test. Preparing, organising and applying a computerised test clearly influenced the practice level. Teachers and students alike had to handle new educational material and get used to using computer systems. Even today, standardised tests like the Stellwerk test are surrounded by controversy concerning the differences between computerised feedback on students' performance and common teachers' evaluations. Students, teachers, school boards and instructors in vocational education must learn how to read such results and interpret them in formative and summative ways. Furthermore, in parallel to the Stellwerk tests, several training platforms were established. One of these platforms is directly connected to the tests themselves and serves as a possible preparation for the assessment. In relation to the new kind of standardised assessment in schools, the Stellwerk test was criticised as a kind of hidden curriculum that endangered instructional quality by encouraging teaching to the test and sanctioning teachers who did not design their teaching according to test contents (Schaller, 2011). The scope of tasks made possible by QTI has also been critically discussed. Item formats like multiple choice questions were not very common in Swiss paper-and-pencil tests (Husfeldt, 2007). Although teachers increasingly got used to these modes of testing, they examined the



results with caution and suspicion. The educational conflicts surrounding the Stellwerk test are symptomatic of the clash between traditional educational practices and the new influence of educational assessment on teaching.

The ÜGK includes no training platforms or teaching materials. In this respect, the evaluation is uncoupled from the instructional level. Nevertheless, an implicit influence can be identified through the information published on students' performance, which is available to teachers and school boards. As yet, it is unclear which instructional conflicts can be expected in the upcoming implementation of the ÜGK. The assumption is that teaching will be more influenced by instruments and arrangements that are closer to the school level than the assessment of the educational system. However, at the same time, given previous experiences with PISA, it seems obvious that this type of system assessment will affect the level of education practice, too.

Implications and conclusion

The educational assessment of students' performance is embedded in a specific educational governmental system and influenced by international trends in educational assessment. By applying an analytical approach to three assessment instruments currently used in Switzerland, we retrace the country's recent history of assessment, progressively complementing existing tests developed locally and from the bottom up with more recent instruments and moving towards a national assessment system of education monitoring.

From a conceptual perspective, we demonstrate that, although the function of newer standardised instruments is conceptualised in a more conscious and clear way than that of tests developed from the bottom up, the functions of actual instrument use tend to be blurred. For assessment policies, this illustrates the importance of defining use



purposes and processes from the beginning and taking into account the possibility of serving several purposes simultaneously. But even with carefully developed and unambiguous concepts, assessment instruments inevitably are subject to recontextualisation processes in policy and practice. Test developers, policy makers and practitioners might as well take these processes into account.

From an evaluation perspective, we demonstrate that the dictum of evidence-based policy leads to the integration of more and more levels of the educational system in the evaluation and information process. It is not always clear which stakeholders hold sovereignty over a given test or who is responsible for the several steps in the educational processes surrounding educational assessment. To address this issue, it is crucial for assessment policies to consider the significance and constraints of various instruments for acting at different levels of the education system and to create an assessment loop that connects the numerous stakeholders in a single dialogue. At the same time, stakeholders do not necessarily share common goals, a collective understanding of a reasonable use of the single instruments or a uniform set of information needs. To be aware of the different positions might already help establish common assessment policies.

From a teaching perspective, we demonstrate that assessment instruments influence instruction both intentionally and unintentionally. With respect to assessment policies, it is important to clarify whether assessment instruments are meant to improve instruction and how they relate to other instruction material and to teachers' professional backgrounds and routines. If assessment instruments are made a part of instruction and designed to contribute to the quality of education, they must be integrated into overarching systems. But still, the actual impact of assessment instruments on instruction and teaching depends on local perceptions of the instruments by school boards and



teachers as well as parents and students. Stakeholders might not neglect this, but find a sensitive approach to handle it.

Following the rudimentary timeline our three instruments under scrutiny represent, the areas of conflict are both persistent and constantly adapting to current developments in education policy and educational assessment influences within Switzerland and beyond.

The confusion of purposes emerged in the early Orientation tests as a lack of coordination between political and instructional aims. Neither the attempts to create a functional divide in the Stellwerk tests nor the limitations of the ÜGK to a mere monitoring purpose fully eliminated this area of conflict. On the contrary, these confusions triggered a reshaping of the traditional differences between the summative and formative purposes of educational assessment both in Switzerland and on an international level. Currently, questions on the asymmetric relationship between the two purposes are combined with questions on appropriate evidence for assessment, assessment quality and the potential dangers and benefits of combinations of purpose (Harlen, 2012).

The influence of test development and the increasing ability to measure competences in more valid, reliable and objective ways opened the way for data aggregation at levels above that of the individual student. The Orientation tests and the ÜGK represent the two poles of this continuum of aggregation and evaluation possibilities. Increased aggregation involves a greater policy focus on schools, regions and nations. As Skedsmo (2011) pointed out for Norwegian evaluation policy, a high evaluative aggregation of assessment data enables the political usage of these data and introduces new inconsistencies between practice and policy.



Using assessment instruments to influence instruction can be seen as a key rationale for implementation in a majority of assessment instruments. The Orientation tests have served as both assessment instruments and instructional material. Furthermore, while the Stellwerk tests are supported by additional training platforms, the ÜGK offers no specific materials for teachers or students. Thus, the monitoring purpose of this assessment instrument inherently involves teaching and instruction. In other words, as Yates and Johnston (2017) state in their examination of teachers' conceptions of assessment in New Zealand, assessment regimes have a strong influence on educational practice and reignite the tension between summative and formative understandings.

Our investigation is meant to instigate a discussion that has not yet fully evolved in Switzerland. The assessment instruments currently in use have not been examined in a systematic or critical manner. Further research could focus on assessment beliefs, practices and policies in the multilevel Swiss education system. The wide variety of policy and practice levels, the multiple relevant stakeholder groups or the different linguistic regions in Switzerland offer possible starting points for such research. Although the political and historical contexts in which these educational assessments are embedded are nation-specific, the broad lines of federalism and such international developments as large-scale assessments and evidence-based policy can also be found in other countries. Thus, as long as both broad political and historical lines and particular backgrounds are taken into account, our analytical approach is transferable to other countries. Comparing assessment tests across different countries and periods from the presented analytical perspectives could support a discussion of the commonalities and differences of assessment policies.



Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributor

Flavian Imlig was a research and teaching assistant in Educational Science at the University of Zurich. In 2017, he changed to the Zurich State Department of Education. His research focuses on the history of education and educational policy. He has been working on the history of teacher training in Switzerland, school reform processes and contemporary initiatives in monitoring educational performance.

Susanne Ender is an academic associate at the Institute for Educational Evaluation, Associated Institute of the University of Zurich. Her work focuses on the development of assessment instruments and educational reporting. She is a PhD candidate at the Institute of Education of the University of Zurich where she analyses the impact of international standardisation and student assessment on the Swiss educational system.

ORCID

Flavian Imlig <http://orcid.org/0000-0002-6305-8088>

Susanne Ender <http://orcid.org/0000-0002-7841-9114>

References

- Ambühl, E., Heller, W., Huldi, M., Oggenfuss, A., Rageth, E., Strittmatter, A., Trier, U. P. (1986). *Primarschule Schweiz: 22 Thesen zur Entwicklung der Primarschule* [Primary school Switzerland: 22 propositions on the development of primary school]. Berne: EDK.
- BD NW (Nidwalden Cantonal Education Department). (2015). *Orientierungsarbeiten: Rahmenbedingungen, Verbindlichkeiten ab Schuljahr 2015/16* [Orientation tests: basic conditions, liabilities from school year 2015/16]. Stans: (n.p.).



- BD SZ (Schwyz Cantonal Education Department). (2015). *Wegleitung für Stellwerk 8 und 9* [Manual for Stellwerk 8 and 9]. Schwyz: (n.p.).
- Berkemeyer, N. (2010). *Die Steuerung des Schulsystems: Theoretische und praktische Explorationen* [Regulation of the school system: Theoretical and practical explorations]. Wiesbaden: Springer VS.
- Berry, R., & Adamson, B. (Eds.). (2011). *Assessment reform in education*. Dordrecht: Springer.
- BFS (Swiss Federal Statistical Office). (1992). *Bildungs mosaik Schweiz* [Education mosaic Switzerland]. Berne: BFS.
- Bieber, T. (2010). Playing the multilevel game in education: The PISA study and the bologna process triggering Swiss harmonization. In K. Martens (Ed.), *Transformation of education policy* (pp. 105–131). Basingstoke: Palgrave Macmillan.
- Bieber, T. (2014). Cooperation or conflict? Education politics in Switzerland after the PISA study and the Bologna process. In K. Martens, P. Knodel, & M. Windzio (Eds.), *Internationalization of education policy. A new constellation of statehood in education?* (pp. 179–201). New York, NY: Palgrave Macmillan.
- Biesta, G. (2010). Why ‘what works’ still won’t work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29(5), 491–503. Retrieved from <https://doi.org/10.1007/s11217-010-9191-x>
- BKD LU (Lucerne Cantonal Department of Education and Culture). (2013). *Orientierungsarbeiten: Merkblatt* [Orientation tests: Information sheet]. Lucerne: (n.p.).
- BKD OW (Obwalden Cantonal Department of Education and Culture). (2013). *Orientierungsarbeiten: Verbindlichkeiten ab Schuljahr 2013/14* [Orientation tests: liabilities from school year 2013/14]. Sarnen: (n.p.).
- BKZ (Conference of Cantonal Ministers of Education of central Switzerland). (2013). *Orientierungsarbeiten im Kontext von Lehrplan und HarmoS* [Orientation tests in the context of Curriculum 21 and ‘HarmoS’]. Lucerne: (n.p.).
- Black, P. (2015). Formative assessment: An optimistic but incomplete vision. *Assessment in education: Principles, policy & practice*, 22(1), 161–177. Retrieved from <https://doi.org/10.1080/0969594X.2014.999643>
- Bölsterli Bardy, K. (2015). *Kompetenzorientierung in Schulbüchern für die Naturwissenschaften* [Competence orientation in natural science schoolbooks]. Wiesbaden: Springer VS.
- Böttcher, W. (2007). Zur Funktion staatlicher ‚Inputs‘ in der dezentralisierten und outputorientierten Steuerung [On the function of state ‘inputs’ for decentralized and output-oriented steering]. In H. Altrichter (Ed.), *Educational Governance: Handlungskoordination und Steuerung im Bildungssystem* (pp. 185–206). Wiesbaden: Springer VS.
- Broadfoot, P., & Black, P. (2010). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7–26. Retrieved from <https://doi.org/10.1080/0969594042000208976>



- Cibulka, J. G. (1990). Educational accountability reforms: Performance information and political power. *Journal of Education Policy*, 5(5), 181–201. Retrieved from <https://doi.org/10.1080/02680939008549071>
- Criblez, L. (2008a). Bildungsforschung und Bildungspolitik oder: von überdauernden Problemen der Grenzziehung: Eine Replik auf Walter Herzog [Education research and education policy: On persistent problems in setting boundaries: A response to Walter Herzog]. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30(1), 153–166.
- Criblez, L. (2008b). Die neue Bildungsverfassung und die Harmonisierung des Bildungswesens [The new education constitution and the harmonisation of the education system]. In L. Criblez (Ed.), *Bildungsraum Schweiz. Historische Entwicklung und aktuelle Herausforderungen* (pp. 277–299). Berne: Haupt.
- Criblez, L. (2008c). Zur Einleitung: Vom Bildungsföderalismus zum Bildungsraum Schweiz [Introduction: From educational federalism to the Swiss educational area]. In L. Criblez (Ed.), *Bildungsraum Schweiz. Historische Entwicklung und aktuelle Herausforderungen* (pp. 9–32). Berne: Haupt.
- Criblez, L., Oelkers, J., Reusser, K., Berner, E., Halbheer, U., & Huber, C. (2009). *Bildungsstandards* [Education standards]. Zug: Klett und Balmer.
- Davies, M. v., Gonzales, E., Kirsch, I., & Yamamoto, K. (Eds.). (2013). *The role of international large-scale assessments: Perspectives from technology, economy, and educational research*. Dordrecht: Springer.
- Davies, H. T. O., Nutley, S. M., & Smith, P. C. (2012). Introducing evidence-based policy and practice in public services. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 1–11). Bristol: Policy Press.
- Drüke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik: Empirische Untersuchungen in neunten und zehnten Klassen* [The culture of tasks in assessments in the subject of mathematics: Empirical studies in 9th and 10th grade]. Wiesbaden: Springer VS.
- EDK (Swiss Conference of Cantonal Ministers of Education). (2011). *Die Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007* [Intercantonal agreement on the harmonisation of compulsory school ('HarmoS-Konkordat') of 14 June 2007]. Berne: EDK.
- EDK. (2013). *Überprüfung der Erreichung der Grundkompetenzen: Konzept* [Evaluation of the achievement of basic competencies: Concept]. Berne: EDK.
- EDK. (2015). *Nationale Bildungsziele für die obligatorische Schule: In vier Fächern zu erreichende Grundkompetenzen* [National education standards for compulsory school: Basic competencies to be achieved in four subjects]. Berne: EDK.
- Fend, H. (2008). *Neue Theorie der Schule: Einführung in das Verstehen von Bildungssystemen* [A new theory of school: Introduction to the understanding of education systems] (2nd ed.). Wiesbaden: Springer VS.
- Frey, P. (2010). Für die Praxis: Orientierungsarbeiten BKZ [For practical application: Orientation tests]. *Schulblatt Nidwalden*, 3, 6–8.



- Goetze, W., Denzler, N., & Wissler, P. (2009). *Evaluation Stellwerk: Kurzbericht* [Evaluation Stellwerk: Short report]. Thalwil: BfB.
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice*, 11(3), 319–330. Retrieved from <https://doi.org/10.1080/0969594042000304618>
- Goldstein, H. (2015). Validity, science and educational measurement. *Assessment in Education: Principles, Policy & Practice*, 22(2), 193–201. Retrieved from <https://doi.org/10.1080/0969594X.2015.1015402>
- Gordon, E. W., & Rajagopalan, K. (Eds.). (2015). *The testing and learning revolution: The future of assessment in education*. New York, NY: Palgrave Macmillan.
- Green, A. (2002). The many faces of lifelong learning: Recent education policy trends in Europe. *Journal of Education Policy*, 17(6), 611–626.
- Green, J. (2011). *Education, professionalism and the quest for accountability: Hitting the target but missing the point*. New York, NY: Routledge.
- Hangartner, J., & Svaton, C. J. (2013). From autonomy to quality management: NPM impacts on school governance in Switzerland. *Journal of Educational Administration and History*, 45(4), 354–369. Retrieved from <https://doi.org/10.1080/00220620.2013.822352>
- Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 87–102). Los Angeles, CA: Sage.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal*, 16(2), 207–223. Retrieved from <https://doi.org/10.1080/09585170500136093>
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379. Retrieved from <https://doi.org/10.1080/0969594970040304>
- Hega, G. M. (2000). Federalism, subsidiarity and education policy in Switzerland. *Regional & Federal Studies*, 10(1), 1–35. Retrieved from <https://doi.org/10.1080/13597560008421107>
- Herzog, W. (2008). Unterwegs zur 08/15-Schule? Wider die Instrumentalisierung der Erziehungswissenschaft durch die Bildungspolitik [On the way to the 'cookie-cutter' school? Against the instrumentalisation of educational science by education policy]. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30(1), 13–31.
- Hovenga, N., & Bos, W. (2009). *Bildungsmonitoring auf der Systemebene* [Education monitoring on the system level]. Düsseldorf: UDiKom.
- Hurrelmann, A., Leibfried, S., Martens, K., & Mayer, P. (2007). The transformation of the golden-age nation state: Findings and perspectives. In A. Hurrelmann, S. Leibfried, K. Martens, & P. Mayer (Eds.), *Transforming the golden-age nation state* (pp. 193–205). Basingstoke: Palgrave Macmillan.
- Husfeldt, V. (2007). *Zum Stand der externen Schulevaluation in Verbindung mit Leistungsmessung: Leistungstests und Schulevaluation in der deutschsprachigen Schweiz und Blick in andere Länder* [The status of external school evaluation



- including performance measurement: Performance tests and school evaluation in German-speaking Switzerland and other countries]. Aarau: PH FHNW.
- Jakobi, A. P., Martens, K., & Wolf, K. D. (2010). Introduction: A governance perspective on education policy. In A. P. Jakobi, K. Martens, & K. D. Wolf (Eds.), *Education in political science. Discovering a neglected field* (pp. 1–20). London: Routledge.
- Jann, W., & Wegrich, K. (2010). Governance und Verwaltungspolitik: Leitbilder und Reformkonzepte [Governance and administrative policy: Guiding principles and reform concepts]. In A. Benz (Ed.), *Governance—Regieren in komplexen Regelsystemen. Eine Einführung* (2nd ed., pp. 175–200). Wiesbaden: Springer VS.
- Jost, D. (1994). *Orientierungsarbeiten Mathematik 5./6. Klasse: Themenschwerpunkte Grundoperationen, Grössen, Brüche, Sachrechnen* [Orientation tests mathematics, 5th/6th grade: The topics of arithmetic, units, fractions, word problems]. Lucerne: BKD LU.
- Klausing, A., & Husfeldt, V. (2015). Verknüpfung von Daten aus Bildungsstatistik und Leistungsmessungen auf Individualebene in der Schweiz [Linking data from education statistics and performance assessments at the individual level in Switzerland]. *Die Deutsche Schule*, 107(4), 352–364.
- Knudson, R. E. (1993). Effects of different instructional tasks on students' narrative writing. *The Journal of Experimental Education*, 61(3), 205–214. Retrieved from <https://doi.org/10.1080/00220973.1993.9943861>
- Koch, S., & Schemmann, M. (2009). Neo-Institutionalismus und Erziehungswissenschaft: Eine einleitende Verhältnisbestimmung [Neo-institutionalism and education science: Introduction to the determination of their relation]. In S. Koch & M. Schemmann (Eds.), *Neo-Institutionalismus in der Erziehungswissenschaft. Grundlegende Texte und empirische Studien* (pp. 7–18). Wiesbaden: Springer VS.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). *Handbook of test development* (2nd ed.). New York, NY: Routledge.
- Larcher, S., & Smit, R. (2011). Unterrichtskompetenz im Berufseinstieg: Mittels 'Mixed Methods' zum Kompetenzmodell [Teaching competency in early career: From 'mixed methods' to a competency model]. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner, & S. Ziegelbauer (Eds.), *Mixed methods in der empirischen Bildungsforschung* (pp. 227–241). Münster: Waxmann.
- Maag Merki, K., & Büeler, X. (2002). Schulautonomie in der Schweiz: Eine Bilanz auf empirischer Basis [Autonomy of schools in Switzerland: A summary on an empirical basis]. In H.-G. Rolff (Ed.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (pp. 131–161). Weinheim: Juventa.
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht: Schülerleistungen messen, bewerten und fördern* [Performance diagnostics in school and instruction: Measuring, evaluating and supporting student performance]. Bad Heilbrunn: Klinkhardt.
- Martens, K. (Ed.). (2010). *Transformation of education policy*. Basingstoke: Palgrave Macmillan.



- Martens, K., Knodel, P., & Windzio, M. (Eds.). (2014). *Internationalization of education policy: A new constellation of statehood in education?* New York, NY: Palgrave Macmillan.
- Martens, K., & Wolf, K. D. (2006). Paradoxien der Neuen Staatsräson: Die Internationalisierung der Bildungspolitik in der EU und der OECD [Paradoxes of the new state: The internationalisation of education policy in the EU and the OECD]. *Zeitschrift für Internationale Beziehungen*, 13(2), 145–176.
- McClelland, D. C. (1973). Testing for competence rather than for ‘intelligence’. *American Psychologist*, 28(1), 1–14.
- Mitchell, D. E., Shippo, D., & Crowson, R. L. (2011). What have we learned about shaping education policy? In D. E. Mitchell, R. L. Crowson, & D. Shippo (Eds.), *Shaping education policy. Power and process* (pp. 286–296). New York, NY: Routledge.
- Newton, P. E. (2017). There is more to educational measurement than measuring: The importance of embracing purpose pluralism. *Educational Measurement: Issues and Practice*, 115. Retrieved from <https://doi.org/10.1111/emip.12146>
- Nussbaum, P., Fischer, S., & Hildbrand, J. (2007). Der Umgang mit Heterogenität in der Schule [Dealing with heterogeneity in school]. In H. Rhyn (Ed.), *Heterogenität, Gerechtigkeit und Exzellenz. Lebenslanges Lernen in der Wissensgesellschaft/OECD/CERI-Regionalseminar für die deutschsprachigen Länder in Nottwil (Schweiz) vom 26. bis 29. September 2005* (pp. 40–50). Innsbruck: StudienVerlag.
- Oelkers, J., & Reusser, K. (2008). *Qualität entwickeln—Standards sichern—mit Differenzen umgehen* [Developing quality, securing standards, handling differences]. Berlin: BMBF.
- Papadopoulos, G. S. (1996). *Die Entwicklung des Bildungswesens von 1960 bis 1990: Der Beitrag der OECD* [The development of the education system from 1960 to 1990: The OECD's contribution]. Frankfurt am Main: Lang. (Original work published 1994).
- Pelgrum, W. J. (1986). *The implemented and attained mathematics curriculum: A comparison of eighteen countries. Second international mathematics study*. Washington, D.C.: Center for Education Statistics.
- Pereyra, M. A., Kotthoff, H.-G., & Cowen, R. (Eds.). (2011). *PISA under examination: Changing knowledge, changing tests and changing schools*. Rotterdam: Sense.
- Rhyn, H. (2009). Evaluation im Bildungsbereich in der Schweiz [Evaluation of education in Switzerland]. In T. Widmer, W. Beywl, & C. Fabian (Eds.), *Evaluation. Ein systematisches Handbuch* (1st ed., pp. 182–192). Wiesbaden: Springer VS.
- Rieder, S. (2005). Leistungs- und Wirkungsmessung in NPM-Projekten: Erfahrungen, Konzepte, Ausblick [Performance and impact measurement in NPM projects: Experiences, concepts, outlook]. In A. Lienhard, A. Ritz, R. Steiner, & A. Ladner (Eds.), *10 Jahre New Public Management in der Schweiz. Bilanz, Irrtümer und Erfolgsfaktoren* (pp. 149–159). Berne: Haupt.
- Roos, M., Wandeler, E., & Mosimann, M. (2013). *Das Übertrittsverfahren Primarschule—Sekundarstufe I des Kantons Luzern: Schlussbericht zur externen*



- Evaluation* [The transition process from primary to secondary school in the canton of Lucerne: Final report of the external evaluation]. Baar: Spectrum3.
- Schaller, R. (2011). ‚Ein Stein im Mosaik der Gesamtbeurteilung‘: Stellwerk [‘A piece of the puzzle of overall assessment’]: Stellwerk]. *ZLV Magazin*, 6, 6–12.
- Sjøberg, S. (2007). PISA and ‘real life challenges’: Mission impossible? In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA. Hält PISA, was es verspricht?* (pp. 203–224). Vienna: Lit.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: Inconcistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5–20. Retrieved from <https://doi.org/10.1007/s11092-010-9110-2>
- Staatskanzlei SG (St. Gallen Cantonal State Office). (2006). *Perspektiven der Volksschule: Bericht der Regierung* [Perspectives on compulsory school: Governmental report]. St. Gallen: (n.p.).
- Sun, Y., & Cheng, L. (2014). Teachers’ grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. Retrieved from <https://doi.org/10.1080/0969594X.2013.768207>
- Sutermeister-Christen, R., Ackermann, U., Aeppli, J., Häcker, T., Luthiger, H., Reinhardt, V., Zutavern, M. (2007). *Lernergebnisse beurteilen und Schülerinnen und Schüler beraten* [Evaluating learning outcomes and advising students]. Lucerne: PHZ.
- Tillmann, K.-J., Dederling, K., Kneuper, D., Kuhlmann, C., & Nessel, I. (2008). PISA als bildungspolitisches Ereignis: Oder: Wie weit trägt das Konzept der ‚evaluationsbasierten Steuerung‘? [PISA as an education policy event, or: How far does the concept of ‘evaluation-based control’ reach?]. In T. Brüsemeister & K.-D. Eubel (Eds.), *Evaluation, Wissen und Nichtwissen* (pp. 117–140). Wiesbaden: Springer VS.
- Tröhler, D. (2013). The OECD and Cold War culture: Thinking historically about PISA. In H.-D. Meyer & A. Benavot (Eds.), *PISA, power, and policy. The emergence of global educational governance* (pp. 141–161). Oxford: Symposium Books.
- Tyack, D., & Tobin, W. (1994). The ‘grammar’ of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31(3), 453–479.
- Tyo, J. (2010). Competency-based education. *The Clearing House*, 52(9), 424–427. Retrieved from <https://doi.org/10.1080/00098655.1979.10113640>
- Uljens, M. (2007). The hidden curriculum of PISA: The promotion of neo-liberal policy by educational assessment. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA. Hält PISA, was es verspricht?* (pp. 295–303). Vienna: Lit.
- Vögeli-Mantovani, U. (1996). *Orientierungsarbeiten: Überlegungen zur klassenübergreifenden Lernerfolgsmessung* [Orientation tests: Reflections on performance assessments across school classes]. Lucerne: ZBS.
- Vögeli-Mantovani, U. (1999). *Mehr fördern, weniger auslesen: Zur Entwicklung der schulischen Beurteilung in der Schweiz* [More support, less selection: On the development of educational assessment in Switzerland]. Aarau: SKBF.
- Weber, H. (2016). Jetzt kommt das Schweizer PISA [Swiss PISA is coming]. *Bildung Schweiz*, 161(2), 12–15.



- Wiater, W. (2013). Kompetenzorientierung des Unterrichts: Alter Wein in neuen Schläuchen? [Competence-oriented teaching: Old wine in new bottles?] *Bildung und Erziehung*, 66(2), 145–161.
- Windzio, M., Knodel, P., & Martens, K. (2014). Reforming education policy after PISA and Bologna: Two logics of governance and reactions. In K. Martens, P. Knodel, & M. Windzio (Eds.), *Internationalization of education policy. A new constellation of statehood in education?* (pp. 247–260). New York, NY: Palgrave Macmillan.
- Wolter, S. C. (2008). Purpose and limits of the education system through indicators. In N. C. Soguel & P. Jaccard (Eds.), *Governance and performance of education systems* (pp. 57–84). Dordrecht: Springer VS.
- Wolter, S. C., & Hof, S. (2014). *Bildungsbericht Schweiz 2014* [Swiss Education Report 2014]. Aarau: SKBF.
- Yates, A., & Johnston, M. (2017). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*. Retrieved from <https://doi.org/10.1080/0969594X.2017.1295020>
- Young, M. D., & Diem, S. (2017). Introduction: Critical approaches to education policy analysis. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis. Moving beyond tradition* (pp. 1–13). Cham: Springer.